**Research on Customer Personality Analysis – Milestone 2**

Sharath Chandra Tummanapally

Department of Data Science, Bellevue University

DSC680 T301 2227 – Applied Data Science

Prof. Catherine Williams

Jun 29, 2022

**Business Problem**: Is it possible for a company to develop its business by researching the personality of its ideal customers?

My main objective is to create a solution or build a model to address this problem of understanding the customer’s attributes like their purchases, income or location and provide a helpful insight to a company that implement this solution.

**Background/History**

An organization can adjust its product based on the target customers from various customer segments by using Customer Personality Analysis. For instance, a business can determine which client segment is most likely to purchase the product and then focus its marketing efforts only on that group of customers, as opposed to spending money to advertise a new product to every customer in its database.

**Data Explanation (Data Prep/Data Dictionary/etc)**

For this step, I identified the unwanted columns and deleted them from data frame as they are not required for our analysis. This refined the dataframe a little. To help us more in analysis, I added new features - ‘Age’, 'NumPurchases' and’ MntTotal' as columns to the dataframe and added their appropriate values. We have added a new column 'Age' to determine the age of the customers and understand the demographics of the customer population and their purchases. I have converted the data frame columns to appropriate datatypes. And finally, I verified the missing data and updated null values to zeros. Thus, we have a refined dataframe.

**Methods**

I am planning to use a Multi Linear Regression model. In our dataset multiple features that are interdependent. Multiple Linear Regression (MLR) is basically indicating that we will be having many features and our output feature.

**Analysis**

When we look at the dataset and our objective of the prediction, we have multiple independent variables impacting a variable - ’MntTotal’. We can say that our Total amount spent on all products (MntTotal) is dependent on Customer’s attribute of Income, Age and Total Number of Purchases made by the customer. Hence, we are choosing a Multi Linear Regression model. Multiple Linear Regression (MLR) is basically indicating that we will be having many features and our output feature. in other words, we have one dependent feature and multiple independent features. Accuracy of this model is around 75%. This better/good accuracy score of our model suggests that our regressive model has fitted well to the existing data. This also suggests that our data is suitable for multi linear regression.

**Conclusion**

Here, we collect data, comprehend and pay attention to what matters in the data, and then build a model to acquire a better understanding and make predictions. Going back to our model I built for this, after analyzing it, I would recommend implementation because the model does perform well, and it has the accuracy around 75%. This also signifies the importance of understanding the customer’s attributes like their purchases, income or location which can further provide a helpful insight to a company that implement this solution. We can improve the model by using it in more datasets that are on the customer purchases across different demographics and refine the model.

**Assumptions**

1. I am assuming that customer spending characteristics are unaffected with the deals and promotions.

2. Customers complains may or may not be linked to the number of web or store purchases made. I assumed that they are.

**Limitations**

1. There are variables like purchase of the fish, fruits & meat are not considered in our analysis. We focused on few variables in our analysis.

2. The data is limited to few customers. More data would have given us more information on Correlation of variables.

3. Loyalty is not a variable but analysis on customers with spending characteristics on essentials like fish, fruits & meat and the deals that can be offered to them can be explored more so we can understand a way to keep them hooked with the company.

**Challenges**

The dataset I had was raw and it was a bit challenging to identify the columns that are dependent and independent. To implement the model I had to add new columns and bring dataset to a effective shape.

**Future Uses/Additional Applications**

This model could work for retail business and e-commerce businesses where understanding the customer’s personality is a huge factor to improve their businesses and grow revenue.

**Implementation Plan**

Step 1: Identifying unwanted columns and deleting them from data frame as they are not required for our analysis.

Step 2: Adding new useful features.

Step 3: Transforming the features.

Step 4: Work on the missing data.

Step 5: Selecting, building, and evaluating a model.

Step 6: Interpreting the results.

**Ethical Assessment**

We have the complete personal data of the customers. There is always a challenge of getting the consent of the customers to use it for this research/project. And this data will be disclosed in this project so there will be a privacy concern as well.